

NAME: Key

Ch. 4 review

ONLINE QUIZ: <http://bcs.whfreeman.com/sta> Go to Ch.4 material, log in (you created this last time), do quiz

MULTIPLE CHOICE:

2. Were the extinctions that occurred in the last ice age more frequent among species of animals with large body sizes? A researcher gathers data on the average body mass (in kilograms) of all species known to have existed at that time. These measurements are values of

- (a) a categorical variable.
- ☒ (b) a quantitative variable.
- (c) an invalid variable.
- (d) a margin of error.

4. An example of a categorical variable is

- (a) the name of the college a student attends.
- (b) a student's weight in kilograms.
- (c) a student's class rank, such as 25th out of 364.
- (d) a student's sex (male or female).
- ☒ (e) Both (a) and (d).

The stock market did well during the 1990s. Here are the percent total returns (change in price plus dividends paid) for the Standard & Poor's 500 stock index:

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Return	31.7	-3.1	30.5	7.6	10.1	1.3	37.6	23.0	33.4	28.6

The next five questions are related to this situation.

5. The median return during this period is

- (a) 5.5
- (b) 20.07
- (c) 23.0
- ☒ (d) 25.8
- (e) 28.6

6. The third quartile of these returns is

- (a) 7.6
- (b) 30.5
- (c) 31.1
- ☒ (d) 31.7
- (e) 33.4

7. The mean return is

- ☒ (a) 20.07
- (b) 20.69
- (c) 22.3
- (d) 25.8
- (e) 33.4

8. The standard deviation of the returns is

- (a) 13.75
- (b) 13.98
- ☒ (c) 14.74
- (d) 20.07
- (e) 25.8

9. You have similar data on returns on common stocks for all years since 1945. To show clearly how returns have changed over time, your best choice of graph is

- (a) a bar graph
- ☒ (b) a line graph
- (c) a pie chart
- (d) a histogram
- (e) a scatterplot

18. A well-drawn histogram should have

- (a) bars all the same width
- (b) no space between bars (unless a class has no observations)
- (c) a clearly marked vertical scale
- ☒ (d) all of these

Here is a table of the undergraduate enrollment at a large state university, broken down by class:

Class	Count of students	Percent of Students
Freshman	8,248	26.8%
Sophomore	8,073	26.2%
Junior	7,001	22.8%
Senior	6,904	22.4%
Non-degree	535	1.7%
Total	30,761	100%

21. To make a correct graph of the distribution of students by class, you could use

- (a) a bar graph.
- (b) a pie chart.
- (c) a histogram.
- (d) all of (a), (b), and (c).
- ☒ (e) (a) or (b), but not (c).

42. For a distribution that is skewed to the right, usually

- ☒ (a) the mean will be larger than the median
- (b) the median will be larger than the mean
- (c) the first quartile will be larger than the third quartile
- (d) the standard deviation will be negative
- (e) the minimum will be larger than the maximum

43. Here is a set of data: 1300, 18, 25, 19, -7, 24. Which observation is the outlier?

- ☒ (a) 1300
- (b) 25
- (c) 19
- (d) -7
- (e) 24

45. You have calculated that the value of the standard deviation for a set of data is -2.5. You can conclude that

- (a) the data are skewed to the left
- (b) the data are skewed to the right
- (c) there is very little variation in the data
- (d) all of the observations in the data set are negative
- ☒ (e) you made an error in your calculations

55. You calculate that the standard deviation of a set of observations is 0. This tells you that

- (a) you made an arithmetic mistake.
- ☒ (b) all the observations have the same value.
- (c) there is no straight-line association.
- (d) the mean must also be 0.
- (e) Both (b) and (d) are true.

57. You are giving a talk about the decay of the American family. You want to show by a graph how the number of divorces has increased between 1970 and 2000. What type of graph should you draw?

- (a) A bar graph.
- (b) A boxplot.
- (c) A histogram.
- ☒ (d) A line graph.
- (e) A pie chart.

69. The five-number summary of a distribution consists of

- (a) mean, median, standard deviation, and two quartiles.
- (b) minimum, maximum, mean, median, and standard deviation.
- ☒ (c) minimum, maximum, median, and two quartiles
- (d) mean, standard deviation, correlation, and two quartiles

72. The standard deviation is a measure of

- (a) the center of a distribution
- ☒ (b) the variability of a distribution
- (c) the association between two variables
- (d) the standardized value of a variable

100. The box in the center of a boxplot marks

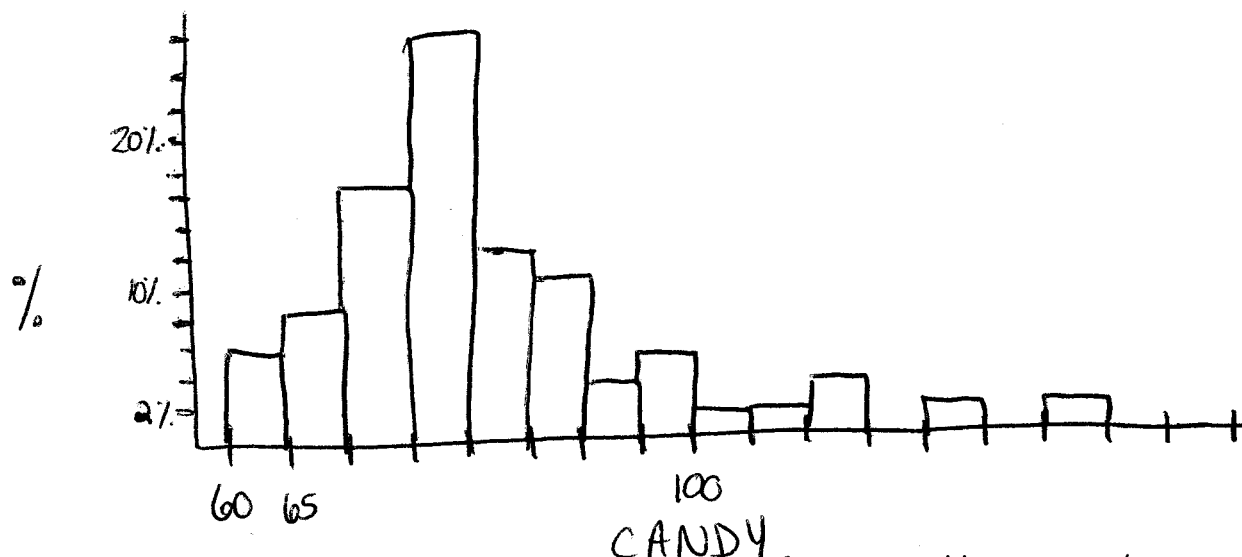
- (a) the full range covered by the data
- ☒ (b) the range covered by the middle half of the data
- (c) the range covered by the middle three-quarters of the data
- (d) the span one standard deviation on each side of the mean
- (e) the span two standard deviations on each side of the mean

FREE RESPONSE:

- 1) The list `CANDY` is a list of the number of pieces of Halloween candy that 50 randomly selected children got on Halloween night.

$n=50$

- a. Create a histogram of the data below, with percent on the y-axis. Be sure to label all your axes!



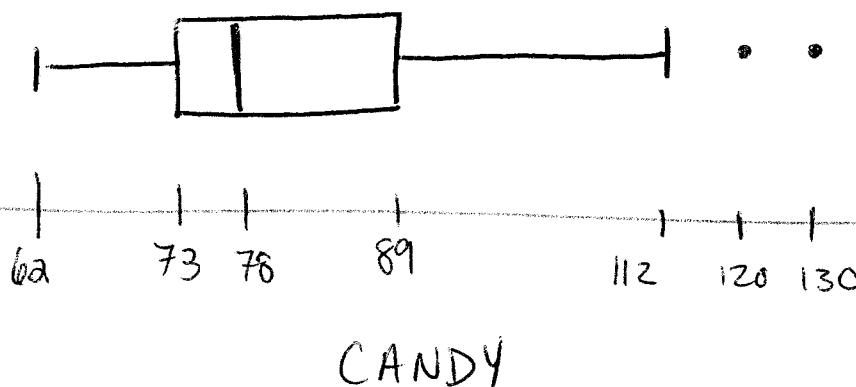
- b. Based on the shape of the distribution, what measure of center would you report (mean or median)?

Median - skewed right

- c. Based on your choice in part ^b what measure of spread would you report (std. dev, or quartiles)?

quartiles - because we chose median

- d. Create a boxplot of the data below.



e. Calculate the following statistics

i. Mean 82.02

ii. Std. dev 14.78

iii. Number of observations 50

iv. Min 62

v. 1st quartile 73

vi. Median 78

vii. 3rd quartile 89

viii. Maximum 130

f. Describe the distribution below (shape, center, spread, etc.)

Shape: right skewed

Center: median of 78

Spread: (62, 112) with outliers @ 120 and 130

g. Test for outliers using the $1.5 * IQR$ test

$$IQR = 16$$

$$1.5(IQR) = 24$$

$$Q_3 + 24 = 113$$

$$Q_1 - 24 = 49$$

outliers
120
130

* anything above 113 or below 49 are outliers

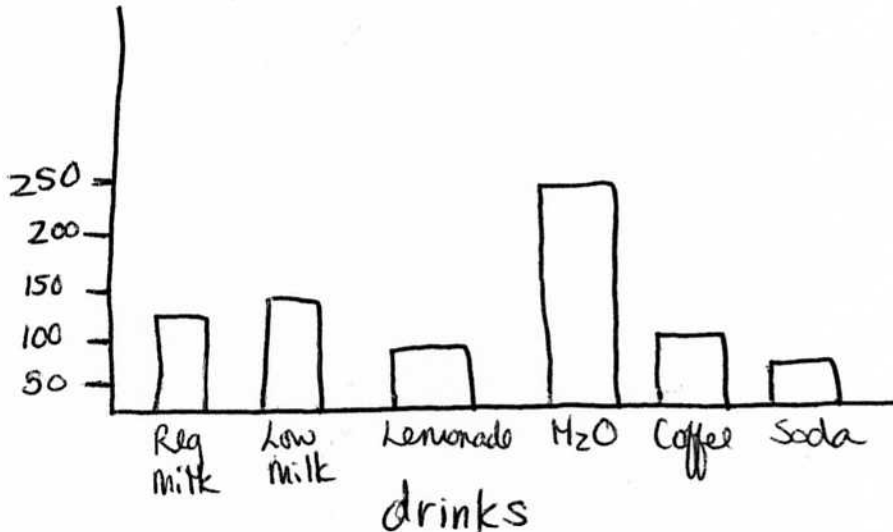
2) Identify the following variables as categorical or quantitative

- a. Number of calculator games on each student's calculator Q
- b. What type of car each student drives (brand of car, Toyota, Honda, etc.) C
- c. GPA Q
- d. Number of mp3 songs each student has on their computer Q
- e. Whether a student is for or against the death penalty C
- f. Amount of time spent studying for the Ch. 4 test Q
- g. Letter grade on the chapter 4 test C
- h. Number of majors offered at a given college Q

- 3) The following are the amount of students who buy each of the different drinks offered in the cafeteria during a 5-day long observational study, at a small school.

		%
✓ Regular Milk	120	15%
✓ Lowfat Milk	136	17%
✓ Lemonade	100	12%
✓ Water	250	31%
✓ Coffee	110	14%
✓ Fountain soda	90	11%
	<u>n = 806</u>	<u>100%</u>

- a. Create a bar graph of the data (use counts on the y-axis)



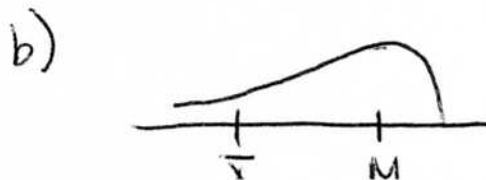
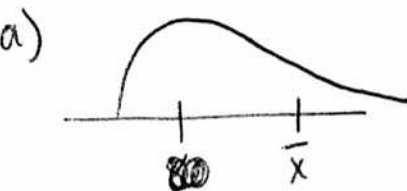
- b. Create a pie chart of the data



drinks

- 4) Given the following info about the mean and median, decide the shape of the distribution of the data, and pick a measure of center to report.

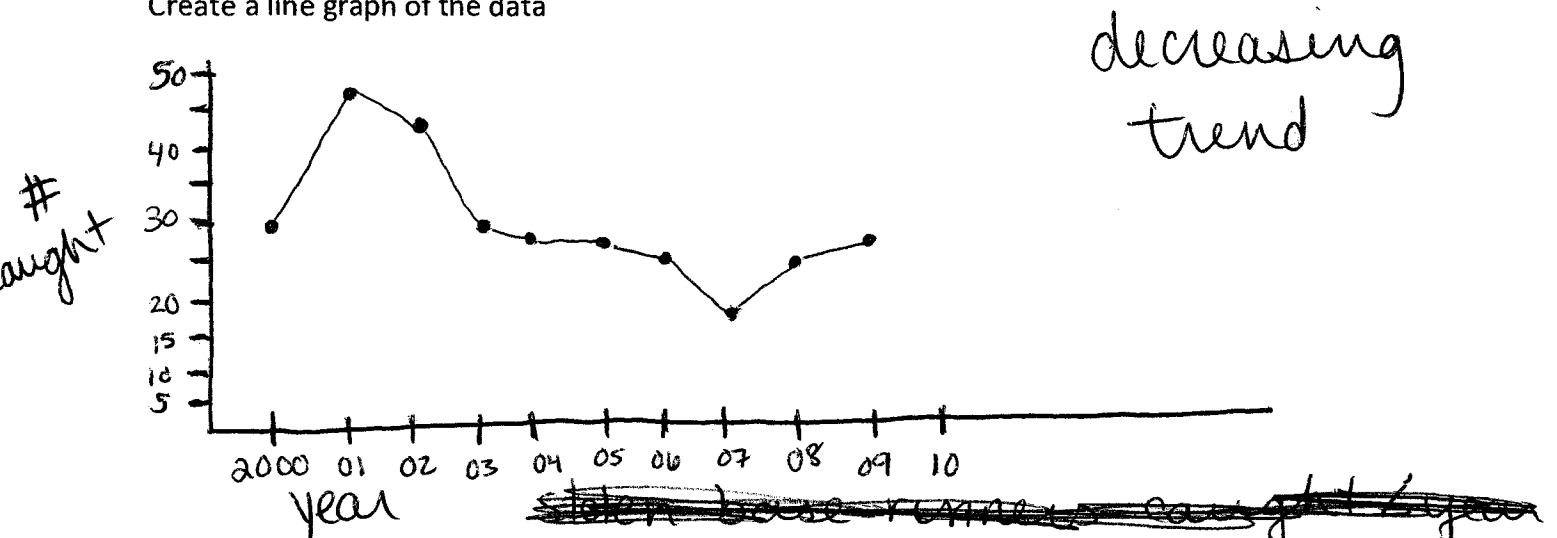
- a. Mean = 170 Median = 80 Shape = rt. skew Center = M
- b. Mean = 100 Median = 180 Shape = left skew Center = M
- c. Mean = 79 Median = 85 Shape = roughly symmetric Center = \bar{X}



- 5) Following are the number of runners caught stealing a base for the Phillies over the past 10 seasons:

YEAR	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
caught stealing	28	25	19	25	27	27	29	43	47	30

Create a line graph of the data



- 6) Using the list `DATA2`, create a stemplot below. You will want to split your stems. How many times you split them is up to you.

```

8 | 2
8 | 5 9
9 | 0 0 1
9 | 5 8 9
10 | 0 2 3 3 4
10 | 6 6 6 7 8 8 8 8 9
11 | 0 1 1 1 2 2 2 3 3 3 3
11 | 5 5 6 6 8 8 8 8 9 9 9 9 9
12 | 0 1 2 2 2 4
12 | 6 6 7 9
13 |

```

- a. Describe the distribution (shape, center, spread)

- Left skewed
- center @ median = 112
- spread = (82, 129)

b. Calculate the following statistics

- i. Mean 110.46
- ii. Std. dev 10.94
- iii. Number of observations 57
- iv. Min 82
- v. 1st quartile 105
- vi. Median 112
- vii. 3rd quartile 119
- viii. Maximum 129

c. Test for outliers, using the mean and standard deviation.

$$\bar{X} + 2(s) = 110.46 + (2 \cdot 10.94) = \cancel{132.34} 132.34$$

$$\bar{X} - 2(s) = 110.46 - (2 \cdot 10.94) = 88.58$$

~~outliers~~

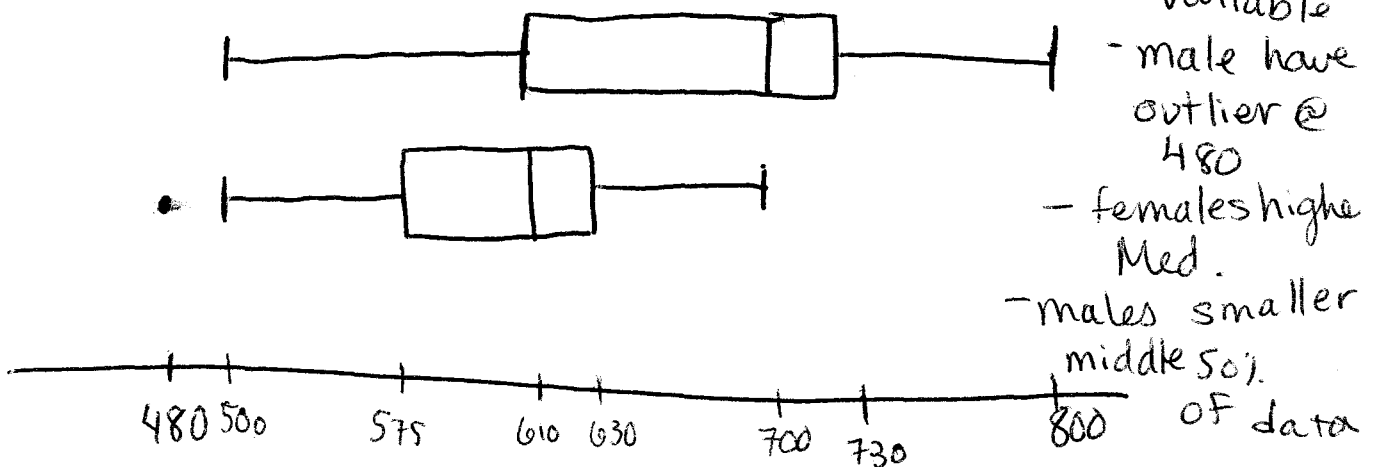
* outliers are #'s above 132.34 and below 88.58

outliers = 82

7) Create side-by-side boxplots for the male and female SAT Math scores (in lists SATMM for males and SATMF for females). Compare the two plots.

Female

male



8) Why are pictograms usually not a good graphical distribution? (the most common reason they aren't good)

- their pictures are distorted
- they don't accurately represent the different categories

9) What is the difference between categorical and quantitative variables?

- usually non-numerical
- put individuals into categories
- numerical
- makes sense to take avg. of them

10) What is a distribution of a variable?

a picture that shows the values of the variable & how often each value occurs

11) What types of values can standard deviation take?

$$S \geq 0$$

Ex: histogram
dotplot
bar graph
etc.

12) If I have a standard deviation of 0, what does this mean?

- all the values of my data are the same

13) What are quartiles? How do they divide up the data?

they divide the data into 4 equal sections (in numerical order)

14) What is the second quartile?

median

15) What percentage of the data is above Q3? 25%.

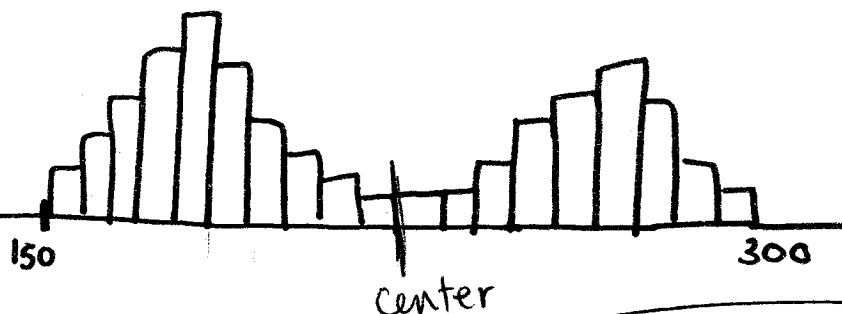
16) What percentage of the data is below Q1? 25%.

17) What percentage of the data is inside the box of a boxplot? 50%.

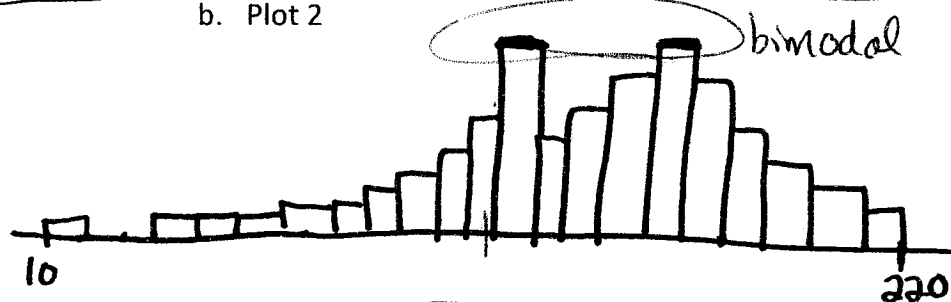
18) Describe the following distributions (shape, center, spread, etc.)

a. Plot 1

- symmetric
- center @ ~ 225
- range (150, 300)
- 2 peaks

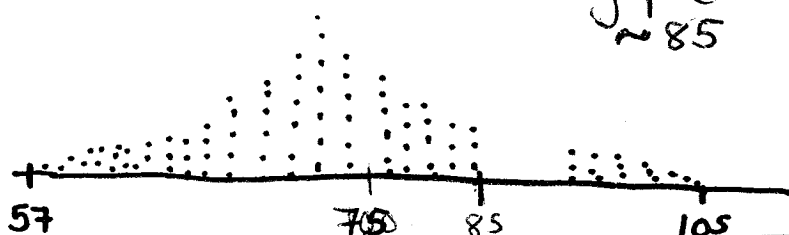


b. Plot 2



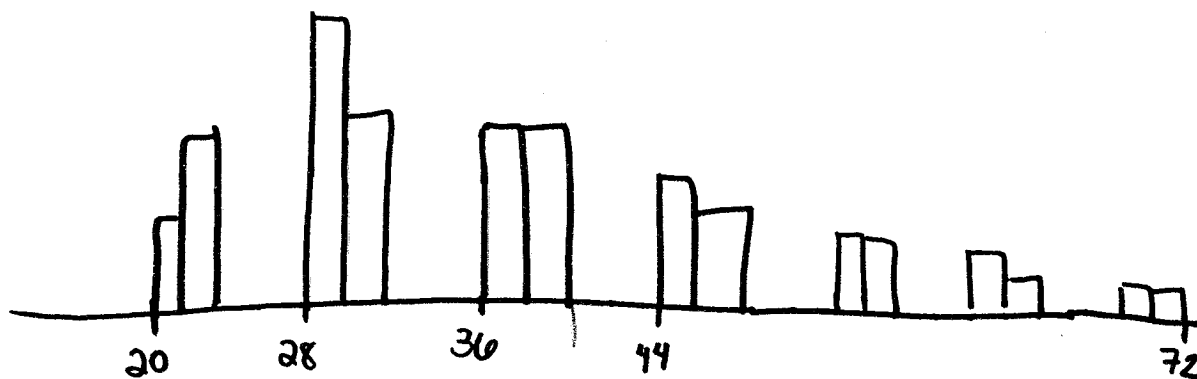
- left skew
- bimodal
- range (10, 220)
- center @ ~ 100

c. Plot 3



- gap @ ~ 85
- roughly symmetric
- range (57, 105)
- center @ ~ 75

d. Plot 4



- granularity
- range (20, 72)
- center @ ~ 40
- shape rt. skewed